AD-A040 630 DECISIONS AND DESIGNS INC MCLEAN VA F/G 5/1
TESTING PROCEDURES IN THE DESIGN OF MANAGMENT SYSTEMS: SOME MET--ETC(U)
FEB 77 R V BROWN, S R WATSON N00014-76-C-0074
UNCLASSIFIED DT/TR-77-1 NL

1 OF 1
ADA
040630

END
DATE
FILMED
7-77

ADA040630

TESTING PROCEDURES IN THE DESIGN OF MANAGEMENT SYSTEMS:
SOME METHODOLOGICAL REFLECTIONS

R. V. Brown
DECISIONS AND DESIGNS INCORPORATED
and UNIVERSITY COLLEGE, LONDON

S.R. Watson
ENGINEERING DEPARTMENT
UNIVERSITY OF CAMBRIDGE

12

DDC
JUN 16 1977
C

ADVANCED DECISION TECHNOLOGY PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

AD No.
DDC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

The objective of the Advanced Decision Technology Program is to develop and transfer to users in the Department of Defense advanced management technologies for decision making. These technologies are based upon research in the areas of decision analysis, the behavioral sciences and interactive computer graphics. The program is sponsored by the Cybernetics Technology Office of the Defense Advanced Research Projects Agency and technical progress is monitored by the Office of Naval Research – Engineering Psychology Programs. Participants in the program are:

Decisions and Designs, Incorporated
The Oregon Research Institute
Perceptronics, Incorporated
Stanford University
The University of Southern California

Inquiries and comments with regard to this report should be addressed to:

Dr. Martin A. Tolcott
Director, Engineering Psychology Programs
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

or

LT COL Roy M. Gulick, USMC
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

The views and conclusions contained in this document are those of the author(s) and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government. This document has been approved for public release with unlimited distribution.

TECHNICAL REPORT 77-1

TESTING PROCEDURES IN THE DESIGN OF MANAGEMENT SYSTEMS:

SOME METHODOLOGICAL REFLECTIONS

R.V. Brown
Decisions and Designs, Incorporated
and University College, London

S.R. Watson
Engineering Department
University of Cambridge

D D C
JUN 16 1977
C

February 1977

Sponsored by

Defense Advanced Research Projects Agency
and
Office of Naval Research
Contracts N00014-76-C-0074 and N00014-74-C-0263

DECISIONS and DESIGNS, INC.
Suite 600, 8400 Westpark Drive
McLean, Virginia 22101
(703) 821-2828

White Section
Buff Section
DISTRIBUTION/AVAILABILITY CODES
SPECIAL
A

SUMMARY

The introduction (or modification) of a management system in an organization is often preceded by an effort to gather data from which it can be evaluated. The data may come from some kind of experiment, a conceptual simulation, or some more informal analysis of relevant past experience. This paper discusses how such alternative testing procedures can themselves be evaluated by paying particular attention to analogous testing paradigms in the more established fields of science and engineering. Decision-aiding systems for naval command and control are used as an illustrative case.

The general principles of scientific sampling are at the logical base of all the approaches discussed. All involve observing how a stimulus (a management system) is associated with some response measure(s) (system performance) in a sample of one or more subjects (decision situations). An inference is then drawn about a target population (i.e., the performance of the system in the situations in which it will operate); this is the system evaluation. Its validity depends on how representative the sample is in terms of the stimuli, the subjects, and the response measures in the sample. Alternative testing approaches differ in how they attempt to achieve each type of sampling representativeness, for example, whether real or surrogate systems are tested or how many different situations are examined.

Classical experimentation, as used in the natural sciences, is the most powerful type of sampling. Stimuli (systems) are actively applied to a systematically selected sample of subjects and the response is observed under conditions of maximal realism in all respects. Such experimentation at its most ambitious is normally too costly and too cumbersome to be applied to the design of a management system. The latter usually involves sequentially scanning a large, complex set of system options, many of them specified only as the design proceeds, and the system is designed for current conditions and therefore is liable to obsolescence. In contrast, because the aim of the natural sciences is to discover time-invariant generalities about a limited set of clear-cut hypotheses, they more readily justify conventional experimentation.

Large-scale conceptual simulation based on judgmental inputs to a computer model is commonly advocated as a cheaper and more manageable alternative to classical experiment, but it suffers from the crucial difficulty of achieving realism, that is, adequate analogy to the real-world setting being simulated. Wargaming is another form of simulation; it

often permits greater realism though its usefulness is limited by difficulties in replication.

A possibly more promising but very different alternative to simulation is prototype testing, as used in engineering; in the context of management systems, this approach would be typified by special purpose fleet exercises for testing naval decision-aiding systems. Involving a few cases thoroughly, prototype testing permits a high degree of realism (but at high cost per observation) to offset its lack of representativeness through small sample size.

A weaker (and much cheaper) variant of prototype testing is the method of clinical observation, as pioneered in medicine. In system design, this method is typified by the use of workshop trials, where a series of historical decisions are examined thoroughly, where decision-aiding systems more or less similar to those being tested were used.

Intuition, or the direct evaluation of system options by expert judgment, is, of course, the loosest and cheapest testing approach and one by no means to be disregarded, given the irreducible difficulties of the others, and it can be combined with one or more of them.

In general, though little appears to have been published on their logical underpinnings, the well-established practices in the engineering design process such as prototype testing appear to offer the most promising models for testing management systems. However, the design of a management system (especially a military management system) differs from conventional engineering design in several respects. The most critical difference is that the setting of ultimate application, namely, war, cannot be adequately replicated during system development, especially in terms of organizational and personal pressures.

The testing approaches considered in this paper differ primarily according to the cost and the accuracy of discrimination among the system options being tested. Typically, of course, the more accurate tests are the more costly. The appropriate choice of test in each instance therefore depends on how important it is to make a definitive discrimination, balanced against the amount of resources available for testing.

As a general rule, the cheaper tests (including intuition in the limit) appear indicated for the last "fine-tuning" modification to a system design, where there is a rich set of potential discriminations among system options to be made and the cost of an error is modest. The more costly tests (including a sequence of increasingly powerful tests from

intuition through workshop trials and simulation to prototype testing) would be reserved for major, clear-cut choices with large costs of errors. An example might be whether the Navy should replace current procedures for contingency planning on aircraft carriers by a computerized pre-programmed decision-making system.

CONTENTS

	Page
SUMMARY	ii
FIGURES	vi
ACKNOWLEDGMENT	vii
1.0 INTRODUCTION	1
2.0 TESTING PROCEDURES AND THEIR PROPERTIES	6
2.1 The Origins of the Testing Problem	6
2.2 Different Testing Procedures	6
2.2.1 Classical experimentation	6
2.2.2 Simulation experiments	8
2.2.3 Prototype testing	9
2.2.4 Clinical observation and workshop trials	10
2.2.5 Intuitive judgment	11
2.3 General Comparison of Testing Procedures	11
2.3.1 Sources of error - a sampling analogy	11
2.3.2 Cost, accuracy, and objectivity	13
3.0 CHOOSING TESTING PROCEDURES FOR STEPS IN THE DESIGN PROCESS	16
3.1 Types of Discrimination in the Design Process - Levels of Generality	16
3.2 The Relation of Test Procedures to Design Discrimination	18
3.3 Sequential Testing	21
REFERENCES	22
DISTRIBUTION LIST	24
DD 1473	32

FIGURES

Figure		Page
2-1	Discriminations in the Design Process Related to Test Procedures	14

ACKNOWLEDGMENT

Many of the ideas in this report derive from work funded by the Office of Naval Research under Contract N00014-74-C-0263 and given a preliminary exposition in Appendix A of Brown et al. (1974). The present report, funded under ARPA Contract N00014-76-C-0074, represents a refinement and development of that material in a different format.

TESTING PROCEDURES IN THE DESIGN OF MANAGEMENT SYSTEMS: SOME METHODOLOGICAL REFLECTIONS

1.0 INTRODUCTION

With the advent of quantitative methods for the analysis of decisions, and sophisticated electronic equipment for transmitting information and making rapid computations, great interest has been developed, particularly in the armed services, in devising systems that will improve decision making. They can vary from fairly simple shipborne computer programs to help a naval commander analyze with great rapidity the complex and important decisions he faces, to developments in the Worldwide Military Command and Control System (WMCCS).

A specific example whose development stimulated the ideas in this paper is a major current research program sponsored by ONR to develop operational decision-aiding systems at task force command level in the Navy.[1] Over a period of some four years (1974-78), this program is intended to lay the methodological foundations and to develop specific guidelines for the design of shipborne systems for use in the planning and execution phases of naval tactical warfare. The final design will involve specifying equipment, procedures, and organization for such decision-aiding systems. Whatever design is ultimately adopted will involve major commitments of resources and often the supplanting of well-established decision processes.

On the way to making such commitments, which are typically sequential, proceeding from broad commitment in principle to fine specification of components, two critical activities are involved: creating alternative designs to consider and discriminating among those alternatives created. The methodological questions raised by this sequence of activities are:

(1) How are alternative designs to be generated (the problem of _invention_)?;

---

[1] R. V. Brown et al., _Decision Analysis as an Element in an Operational Decision Aiding System_, Technical Report 74-2 (McLean, Va.: Decisions and Designs, Incorporated, September, 1974); Brown et al., _Decision Analysis as an Element in an Operational Decision Aiding System (Phase II)_, Technical Report 75-13 (McLean, Va.: Decisions and Designs, Incorporated, November, 1975); and C. R. Peterson et al., _Decision Analysis as an Element in an Operational Decision Aiding System (Phase III)_, Technical Report 76-11 (McLean, Va.: Decisions and Designs, Incorporated, October, 1976).

(2) How are proposed designs to be compared in light of available evidence (the problem of evaluation)?; and

(3) How is evidence to be gathered in preparation for evaluation (the problem of testing)?

It is clear that these are not problems peculiar to the design of a decision-aiding system or even management support systems generally (including accounting, control, communication and data-gathering systems). Systems engineers have extensive experience in system design and have discussed appropriate design methodologies;[2] the process of designing a specific machine is essentially the same.[3] But it is clear that the appropriate response to these methodological questions depends upon what is being designed.

We shall say only a little about the problem of invention since it seems to be least affected by the nature of the object of design, and no systematic way is known to handle it. Lucas[4] gives guidelines for creative design, but even these cannot be said to be general.

While this problem of generating alternative designs or system choices is of obvious importance in improving technological performances, it does not prevent at least some technical advance. The possible existence of better alternatives did not prevent NASA engineers from coming up with a superb design for Apollo. In any case, however important the inventive part of design is, it is not our task in this paper to seek a generalized methodology for it, if any exists. It may forever remain in the province of art, much like the generation of scientific hypotheses.

The problem of evaluation, however, is more susceptible to methodological scrutiny since the effective features of design are available to measurement and quantification. For example, in designing a machine to make bolts to a given tolerance, the variables to evaluate are clear, namely, the dimensions of the bolts within the given tolerance, the cost of the machine, and, perhaps, the productivity and durability of the machine. More ambiguous are the appropriate evaluation

---

[2] Wilton P. Chase, Management of Systems Engineering (New York: John Wiley, 1974).

[3] G. L. Glegg, The Design of Design (Cambridge, Eng.: Cambridge Univ. Press, 1969).

[4] H. C. Lucas, Jr., Toward Creative Systems Design (New York: Columbia Univ. Press, 1974).

measures for a more complex system, for example, in the design of a telecommunications network. Here, as in all design work, cost is important. But whether and how specifically the quality of service to the user should be included among the performance variables is not at all clear. In the design of a decision-aiding system, the problem becomes much more acute. Such a system is supposed to improve the quality of decision-making, by which we mean that the chances of attaining the goals of the decision-making body are increased. In a naval context, these may be improving the chances of winning wars or, ultimately, more effectively discouraging the enemy from attacking. Although it will sometimes be clear that a happy outcome of a decision can be traced to a new decision-aiding technology, it will not always be the case. The same decision might have been taken without the technology, or another decision might have been taken with an even happier outcome.[5] Thus, evaluating a decision-aiding system or a component of one raises acute problems in measuring performance even if the evaluatory variables have been specified.

Once the evaluatory variables have been measured, there is the further methodological problem of combining them into a single index of performance. The theory of multi-attributed utility measurement[6] successfully deals with this issue; it has been applied to the evaluation of military weapon systems.[7] Work along the same lines has also proceeded in the area of evaluating operational decision aids where multiple performance measures are assigned to each alternative by experts and weighted according to their assessed importance for the purpose at hand.[8] The methodology of evaluation per se, therefore, does not call for special attention in this paper.

---

[5] S. R. Watson and R. V. Brown, Issues in the Value of Decision Analysis, Technical Report 75-9 (McLean, Va.: Decisions and Designs, Incorporated, October, 1975).

[6] R. L. Keeney and H. Raiffa, Analysis of Decisions with Multiple Objectives (New York: John Wiley, 1976).

[7] M. L. Hays, M. F. O'Connor, and C. R. Peterson, An Application of Multi-Attribute Utility Theory: Design-to-Cost Evaluation of the U.S. Navy's Electronic Warfare System, Technical Report DT/TR 75-3 (McLean, Va.: Decisions and Designs, Incorporated, October, 1975); and J. O. Chinnis, C. W. Kelly, III, R. D. Minkler, and M. F. O'Connor, Single Channel Ground and Airborne Radio System (SINCGARS) Evaluation Model, Technical Report DT/TR 75-2 (McLean, Va.: Decisions and Designs, Incorporated, September, 1975).

[8] Brown et al. (1975).

On the contrary, the testing phase in the design process, concerned with gathering the data on which evaluation is based, has a quite ambiguous methodological status. At its most abstract level, the logic is well understood. A choice between alternative testing procedures can be analyzed according to the "value-of-information" paradigms of personalist decision analysis.[9] However, the prior clarification of considerations to be modeled in such an analysis, which includes identifying promising options for information gathering, is not well understood.

In looking for promising information, that is, testing options for system design, it is instructive to examine analogies in the procedures of scientific research (just as we have earlier drawn on analogies from general engineering design.) Different hypotheses are invented, testing procedures (such as experimentation) are used to provide data bearing on the validity of the hypotheses, and the hypotheses are evaluated.

In the testing phase, with which we are now concerned, well developed paradigms based on sampling theory and the classical theory of experiments abound in the literature and are often urged as the model for system design. In considering adapting that model to the needs of system design and specifically to the needs of military decision-aiding systems design, attention must be paid to differences of approach between scientific discovery and engineering design. As Glegg puts it:

> The Engineering Scientist and the Natural Scientist travel the same road but sometimes in opposite directions. The Engineer goes from the abstract to the concrete; other scientists from the concrete to the abstract. The Astronomer takes most careful and exact measurements of a planet and then deduces its future position and movements in the form of abstract mathematical formulae. The Engineer's work is the converse of this. He invents with his imagination and he builds with his hands.[10]

The transparent importance of somehow bringing persuasive empirical evidence to bear on design proposals before major resources have been committed to them, coupled with the

---

[9] H. Raiffa, Decision Analysis (Reading, Mass.: Addison-Wesley, 1968).

[10] G. L. Glegg, The Science of Design (Cambridge, Eng.: Cambridge Univ. Press, 1973), p. I.

prestige that attaches to the classical methods of science, has encouraged many military researchers to advocate testing procedures for system designs based on the approach of classical experimentation. Given the fundamental differences between scientific and engineering approaches, we believe, on the contrary, that the paradigms of the engineering design process, though less well developed in the theoretical literature, may prove to be a more promising model for the testing of decision-aiding systems.[11]

The main task of the paper, then, is to explore in a tentative and discursive way the relative merits of alternative approaches to testing different types of options arising in the design of management systems in general and military decision-aiding systems in particular. We shall use naval systems as an example since it is these that have suggested this inquiry. We shall compare methods for evaluating the performance of engineering systems and methods for evaluating scientific hypotheses, being the main, well-established models available. For descriptive convenience, we shall consider the design process as depending on a sequence of discriminations, that is, statements or hypotheses about the relative appropriateness of alternative design options as they unfold over time.

---

[11] D. L. Marples, "The Decisions of Engineering Design," Management Transactions (Institute of Engineering Designers) (1961): 1-16.

2.0 TESTING PROCEDURES AND THEIR PROPERTIES

2.1 The Origins of the Testing Problem

Unlike the problems of invention and evaluation, the problem of how to test alternative designs has features highly dependent on the object of design. Thus, in designing a bracket to support a given load, we can make a few brackets and see which is best very easily. However, the designers of Concorde had to distinguish between many design options by making mathematical models and using them to test which design was better; they could not afford to build several different aircraft just to test the tail design! If discrimination between designs is based on a model of a system (whether conceptual or physical prototype) rather than on the system itself, then the inference that one design is superior to others becomes subject to substantial uncertainty. In such cases it is more accurate to speak of pretesting rather than testing the design.

Moreover, many systems work in an inherently variable environment so that, even if a comparison of these systems could be made by using the actual fabricated designs, the observation that system A performed better than system B in one situation need not imply that it would do so in another. Thus, a naval decision-aiding system well-suited to open-sea warfare might be quite ill-suited to an amphibious landing. It is the presence of uncertainty both in system performance and in inference that makes the choice of testing procedures important for complex systems, particularly those constructed to do something as difficult to assess as improving decision making.

2.2 Different Testing Procedures

As we have seen, testing involves problems of dependency upon the object of design, the environment in which the design is used, and the interrelated factors of accuracy and cost. Accordingly, each approach to testing has distinctive features which address these problems in different ways. We shall now describe some specific testing procedures and later (in Section 2.3) consider the ways in which they differ in terms of accuracy and cost.

2.2.1 Classical experimentation - The classical experiment is the testing procedure most characteristic of the "scientific method." It is the method usually chosen for testing hypotheses about accessible and measurable material, for example, the response of agricultural crops to alternative

treatments.[1] A typical hypothesis would be that a certain stimulus or treatment (e.g., a fertilizer) applied to a certain subject or subject matter (e.g., a strain of wheat) with certain disturbance factors held constant (e.g., soil and climate) will produce a performance measure or response (e.g., grain yield) with certain statistical properties (e.g., the yield is greater on average than with no fertilizer). A sampling test of the hypothesis would essentially consist of measuring the response of the material to the treatment in a sequence of instances under varied factor conditions. If, in addition, the treatments are deliberately applied (as opposed to simply being observed as in other samples), we have an experiment. For example, the fertilizer would be applied to (or withheld from) wheat in a variety of plots (soil factor), and the yield in each case would be measured.

An experiment can be regarded as a special kind of stratified sample, with one layer of stratification being the treatment and other strata corresponding to disturbance. The design of experiments (as of stratified samples) is concerned primarily with selecting experimental units for efficiency of inference, for instance, to discount the effect on response of factors other than the treatment itself. Sophisticated blocking schemes such as Latin squares are used to balance the needs of inference generality and economy of effort in the assignment of observations to blocks (or factor strata).

A critical requirement of an effective experiment is that the stimuli, subject matter, disturbance factors, and response measures used in the observations closely reproduce the world to which the hypothesis is intended to apply, as they do in the agricultural example. There is no reason in principle why alternative system designs should not be tested by classical experimentation, provided that this reproducibility requirement is met.

Sometimes it is, when cheaply produced engineering hardware is involved. A hypothesis is set up (e.g., that carbon fiber is better than steel for turbine blades), experiments are constructed to test that hypothesis (real turbines are constructed and tried with blades of both materials), the results of the experiments are noted (the thrusts and lifetimes of both engines are noted), and an inference is drawn (steel is better than carbon fiber). But

---

[1] R. A. Fisher, Design of Experiments (1951; rpt. New York: Hafner Press, 1974), and see D. R. Cox, Planning of Experiments (New York: John Wiley, 1958).

in the design of unique systems, this procedure ceases to be feasible. We cannot build two alternative freeway systems for a city to see which performs better.

Much less can we repeat military engagements with and without a given decision-aiding system to determine whether it is an improvement in that kind of engagement. In this case, the stimuli to be evaluated are man-machine systems, and the subject matter is tactical situations requiring a decision. Although the objective is still to evaluate the differential effect of different systems (or different elements of systems), reproducibility is immensely more difficult to achieve. The stimuli may be far from being in a well-defined and therefore reproducible form (unlike a fertilizer compound), and disturbance factors may be vastly more numerous and equivocally defined than soil and climate in agriculture, the response measures may be difficult to define (unlike the yield of a wheat crop), and the subject matter may be impossible to replicate in an experimental setting.

2.2.2 Simulation experiments - An increasingly common alternative to classical experimentation is to test a hypothesis about a system with experiments not on the real system but on a conceptual simulation of that system. For example, the conjecture that a steel works functions more efficiently if a new process control system is introduced could be investigated by constructing a computer-based simulation of the steel works and then seeing how the model reacts to the new process control system. The validity of inferences from such an experiment is clearly limited by how close the analogy is between the simulated and the real system. It is usually necessary to expend a great deal of time and money in getting an adequate simulation model, though less, of course, than doing experiments on the real thing.

Basically, there are two ways of generating a simulation trial for testing purposes, that is, running a subject once through the conceptual experiment. One way is the classical Monte Carlo simulation approach, in which the environment is characterized by a probabilistic model with judgmentally preassigned structure and parameters. Any particular simulation trial is then randomly generated by standard Monte Carlo means. A set of such trials represents a random sample from the probability distribution of outcomes implied by the model (with its input judgments), and it is repeated for each system option. In this way a sample estimate of the expected performance of each option is obtained. However, it is still only an estimate of performance in the simulated system, which may itself bear only weak comparison with the real system.

To obtain an adequate representation of an extremely complex environment (such as that of a military engagement) requires a massive modeling effort, yet it is not clear that any feasible amount of effort would be able to produce a prestructured environment that captures the essential features of the actual environment. In fact, some very important factors, such as the effects of bureaucratic pressures in the real world, are virtually impossible to simulate in the laboratory. The accuracy of the model is entirely dependent on the quality of the judgment on which it is based.

There is no consensus among researchers in the field on how appropriate this type of simulation is for social research in general. Based on experience with one multi-million dollar simulation experiment at a major university over the period 1962-1968, the two major participants formed quite different conclusions. The main technical consultant emerged favorably inclined toward simulation experiments of this kind, but the project manager was not pleased by it and discouraged future efforts of this kind. We are inclined to share his negative view.

A second form of simulation experiment, often used for training military officers, is <u>war-gaming</u>. Here, some or all of the environmental response to the stimulus (decision-aiding system) is generated by military experts who serve as an environmental "surrogate" for such factors as enemy reaction or weather that might affect the performance of the system. Contingencies do not, therefore, have to be anticipated ahead of the exercise, as in Monte Carlo simulation. A sophisticated form of war game can be obtained with the use of the step-through simulation approach, where probability distributions (rather than single responses) are supplied by the environmental surrogates as called for, and then they are randomly sampled.[2] The output of step-through is indistinguishable from regular Monte Carlo. However, like other types of war gaming, it is cheaper and less liable to incorrect structure than conventional simulation and thus appears more appropriate in general for present purposes.

2.2.3 <u>Prototype testing</u> - A testing approach diametrically opposed to simulation but exploiting different features of classical experimentation is prototype testing. In this approach, stimulus and subject are replicated as exactly as possible (and typically expensively) but with few iterations

---

[2] Brown et al., (1974).

and are possibly unrepresentative of disturbance factors. In engineering, the stimulus might be an exact replication of the target stimulus (e.g., a wing of the design to be tested), but the subject might be only approximated (a wind tunnel rather than an actual flight test). Similarly, a fleet exercise might use a very close approximation to the actual decision-aiding system (stimulus) but much less exactly replicate the ultimate war setting (subject). Other applications of the prototype testing approach include test marketing in business and medical testing of new drugs on animals.

This prototype testing approach usually does not for reasons of cost permit many trials or control of their selection (contrasted with computer simulation). It seems better adapted to the evaluation of well-defined designs than generalized principles.

2.2.4 <u>Clinical observation and workshop trials</u> - A further testing approach similar to prototype testing, more feasible, but weaker, is the method of clinical observation used in medical research. The difficulty with experimenting on human subjects encourages medical researchers to make judgments on the basis of case histories that have come their way. Observation of case histories of patients with a given condition lends support or opposition to hypotheses both about the kinds of circumstance in which the condition arises and about the most appropriate treatments for it. This procedure is similar to the development program on an engineering design. A new airplane is produced, and, in the light of experience with that plane, design modifications are made to produce a Mark II model which performs better, and so on.

A modification of clinical observation, adapted to decision-aiding systems is the method of "workshop trials." The case studies are reports of real decisions made in specific decision situations in specific scenarios and based upon whatever decision aids (system option) happened to have been used. The case decisions are re-analyzed in the laboratory and extended imaginatively by alternative decision-aiding options. The value of such analysis depends upon the realism with which interaction of the real world situation and the proposed option is recreated. The participation of qualified observers, perhaps even the decision maker involved in the original case study, is required.

An actual instance illustrates this approach. To test the broad principle of incorporating preprogrammed decision analysis into contingency planning for tactical naval operations and also to suggest promising directions for more specific elaboration, a specific tactical incident

was chosen as a subject: the bombing of a power plant in Haiphong during the Vietnam war.[3] The way the decisions leading up to the bombing were actually made were described by the responsible task force commander. He was interrogated at length on the value and drawbacks of the decision-making process he used and on his perception of the potential implications of having had available other facilities such as decision aids of the type being considered. His responses led to transferable insights on where such aids would be most promising and what form they should take.

This "clinical" approach can be considered as a very special case of uncontrolled sampling, where, unlike experimentation, no attempt is made to impose a stimulus on the subject observed. The more usual kind of uncontrolled sampling would involve random or systematic observations drawn from a population of decision situations, where the variations in stimulus (system option) occur as they will. This approach appears quite inappropriate in our situation, where the stimuli under investigation do not spontaneously happen since they are design proposals which do not pre-exist in any population that can be sampled.

2.2.5 Intuitive judgment - Still further in the direction of informality is the possibility of making a discrimination among design options by using intuitive judgment based on personal experience. As we have noted before, this method can give spectacularly incorrect inferences, like Lord Rutherford's claim that nuclear power was technologically inaccessible. It does, however, have the advantage of low cost, which suggests that little is to be lost by using it in conjunction with other more powerful approaches.

2.3 General Comparison of Testing Procedures

2.3.1 Sources of error - a sampling analogy - We have not given an exhaustive list of testing methods in the last section, but the ones we have given span the range of promising alternatives. These methods can all be considered to be sampling inquiries. The target population of subject matter is, say, all military engagements by a certain naval task force over the next ten years. The purpose of the design testing inquiry is to estimate the average response (some summary performance measure) in this population of alternative stimuli (decision-aiding systems). Since the target population cannot usually be exactly and exhaustively measured, an attempt is made to take realistic subjects under representative conditions and subject them to the stimuli of interest. Finally, an inference is made about

---

[3] Brown et al. (1974).

stimulus-response association in the target population by analogy to the sample as measured. The strength of this analogy and, therefore, the accuracy of the test depends on how closely the sample corresponds to the target in the following respects:

(1) The sampling subject (plot in an agricultural experiment) may be more or less realistic, that is, similar to the target subjects (e.g., real military engagements). For example, my conception of a ship's environment in the heat of battle, an admiral's conception, a computer simulation of it, or mock engagement in a fleet exercise may or may not correspond to an actual engagement in terms of external threats and psychological and organizational pressures that bear on the effectiveness of a decision-aiding system.

(2) The treatment or stimulus may differ from the one to be tested. For example, the equipment used in a fleet exercise to communicate information needed for a decision system may be only a primitive form of what would really be used.

(3) The response or performance measured on the sample subjects may be an imperfect surrogate for the real measure. For example, the task force commander's subjective feelings of satisfaction or the number of aircraft shot down may be used in place of the more elusive but more relevant propensity to win wars.

(4) The sample may be more or less representative of the target population in terms of disturbance factors because of sample size or method of selection. Properly stratified random sample designs or controlled experiments are among the more powerful (but, often more difficult) techniques for ensuring sample representativeness, followed by opportunistic quota sampling (e.g., making sure at least some instances of the major types of scenarios are included); and, finally, completely unconstrained catch-as-catch-can accumulation of instances (e.g., interviewing admirals who happen to be available).

(5) The inference from the sample to the target population may be done in a more or less sophisticated way, from direct informal judgment through classical statistical inference to Bayesian updating and other personalist approaches.

These various testing procedures differ along one or more of these dimensions, each of which introduces a

different source of error in estimating the target values.[4] Thus: classical experimentation (where feasible) scores high on all components of accuracy; computer simulation controls sampling error (because of the large random sample possible) but not other kinds of errors (because the sample may be from the wrong population); prototype testing scores high on subject stimulus and response realism (because of good insights from the cases examined) but low on sample representativeness (because of the limited number and choice of cases); clinical testing is intermediate between prototype testing and simulation on realism and representativeness; intuition can be considered as informal inference on an unconstrained sample of whatever one's experience has been, and scores low in all respects.

2.3.2 Cost, accuracy, and objectivity - The two main considerations affecting a choice among testing procedures at any stage in the design of a decision-aiding system are accuracy and cost. The purpose of inferential procedures in the face of uncertainty is to discover the truth; the probability of being correct (or of making small errors) in such an inference is a measure of the accuracy, or power, of such a test. From an examination of sources of error in the previous section, we can assert that accuracy generally increases as we move from intuition to controlled experiments (see Figure 1). It is also true that accuracy can be increased at a cost for any particular testing procedure. The judgment in choosing a testing procedure and in choosing the appropriate mode for the procedures lies in trading off accuracy against cost.

Roughly speaking, as we move from controlled experiment to intuition, the cost of making a discrimination decreases, as indicated in Figure 1. It is cheaper to select steel for turbine blades on the basis of a hunch than to build several test engines to see whether carbon fiber would be better. There are, of course, great variations in cost possible in the application of a testing method to a particular design choice. The cost of random sampling in experiments, for example, is generally an increasing function of the sample size.

There are other attributes of test procedures, usually less important, which may affect our choice, not least of which is the degree to which the test procedure is

---

[4] The technique of decomposed error assessment (DEA) is available to analyze and aggregate quantitatively such sources of error. See R. V. Brown, Research and the Credibility of Estimates (Cambridge, Mass.: Harvard Business School, 1969).

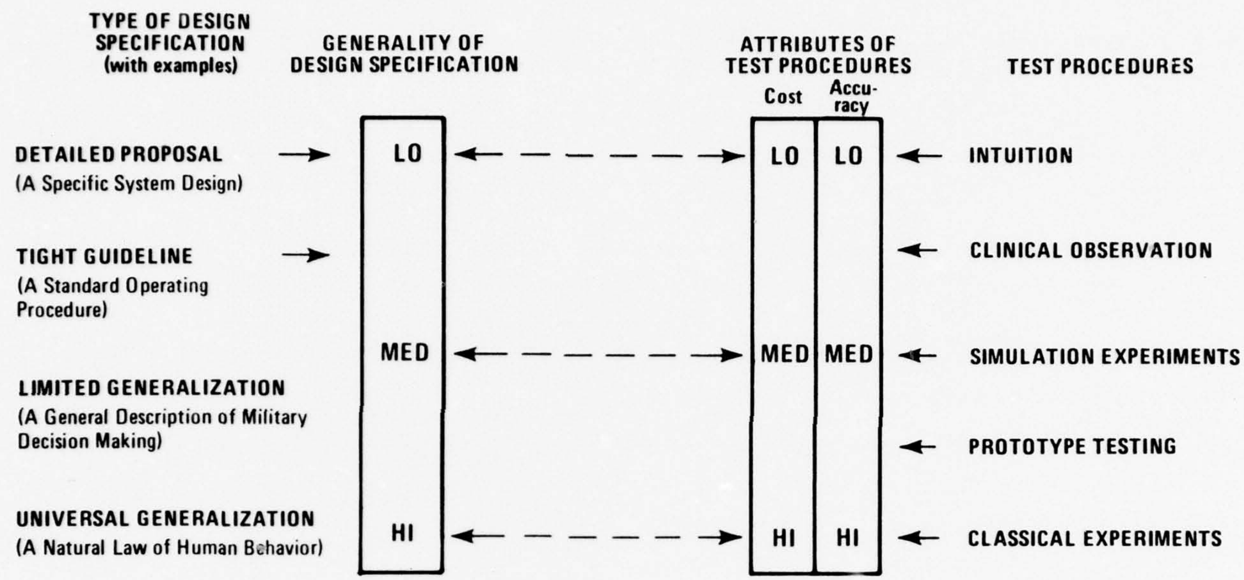

Figure 2-1

DISCRIMINATIONS IN THE DESIGN PROCESS RELATED TO TEST PROCEDURES

objective or independent of the evaluator. This is particularly important if the results of the testing procedure are to be accepted by others or if the competence or objectivity of the tester is in question. Of course, if the tester is himself the sole decision maker and no persuasion of others about the validity of the test is called for, this feature loses in importance.

3.0 CHOOSING TESTING PROCEDURES FOR STEPS IN THE DESIGN PROCESS

3.1 Types of Discrimination in the Design Process - Levels of Generality

Technical discriminations to be made in the process of designing a decision-aiding system differ on many respects bearing on the most appropriate procedure for testing them. They also differ with respect to the stage in the design process (early coarse determinations such as whether to use decision analysis at all are of a different kind from later fine-tuning like the choice of computer hardware), the part of the system being specified (organization, equipment, procedures), and the function to which the system will be put (military versus civilian, operational versus procurement). But perhaps the most helpful classification of these discriminations is according to the generality of the proposition being tested. The left-hand column of Figure 1 is an attempt to capture this classification.

At the most general are statements about human abilities or social structures (e.g., "the untrained assessor tends to assess probability distributions which are tighter than the evidence warrants" or "under pressure, humans favor those decision-making processes which they learned first"). Conceptually, these could, in the limit, have the force of natural laws, but the social sciences have not yielded many such laws, and, in practice, one works with much weaker statements.

Slightly less general are assertions concerning decision-making techniques and processes which refer to a specific setting or culture (e.g., "in the execution of modern warfare strike decisions, the dominant criteria for performance of decision-aiding systems are speed of response and conceptual completeness," or "the technique of 'staged-tree' analysis demands more skill and time at the current state-of-the-art than the direct value approach").

At the next more specific level are recommendations to follow tight guidelines in developing or operating a system along the lines of standard operating procedures common in the Navy (e.g., "use preprogrammed decision aids for contingent decisions where the expected number of occurrences is at least one, and where..." or "never use the 'step-through Monte Carlo' method for analyzing decisions at the

execution phase in tactical warfare").[1] Generalizations at this level might form a primer for systems designers at the detailed level.

A still less general kind of specification concerns the use of a particular system of hardware and software ready to install on board a ship, (e.g., "component X is preferred to component Y for System Z" or "the triangular form of contingent analysis display should be replaced by a device to display n-valued hypotheses in this particular case"). The most specific kind of statement, of course, is a recommendation to install a completely specified system on specified ships at a specified time.

The design process for management systems involves statements and discriminations at all levels of generality. Typically, the generality of the discriminations that need to be made, starting with the choice of a general idea and ending with the elaboration of a finely specified product, decreases as the design proceeds. In the design of Concorde, for example, generalizations about the natural world, namely, the laws of physics, were used to support engineering doctrine which, in its turn, underpins standard engineering methods. Although the design engineers might well have challenged some standard engineering methods and even questioned some tenets of engineering doctrine, they were very unlikely to have questioned the laws of physics. As the design process unfolds, discrimination and decision become more significant and happen at a level of increasingly lower generality.

Analogously, in the design of decision-aiding systems, generalizations about man and societies of men support hypotheses about particular processes in decision-making procedures which are used by the systems designer to produce a particular decision-aiding system. Once again, at the high levels of generality, there is little for the system designer to do in discrimination, unless he wishes to engage in scientific research as well. He must accept prevailing assumptions about aspects of human behavior and even about the nature of decision-making processes in the military. Although he will have a good deal of latitude in the specification of standard operating procedures, most of his effort will be in discrimination at the detailed level of specification of particular system components and their interrelations.

---

[1]This level of generalization is roughly that used to match analytic options to decision situations in R. V. Brown and J. W. Ulvila, Selecting Analytic Approaches for Decision Situations: A Matching of Taxonomies, Technical Report 76-10 (McLean, Va.: Decisions and Designs, Incorporated, October, 1976).

3.2 The Relation of Test Procedures to Design Discrimination

In this section, we attempt to justify the hypothesis that a management system should be tested by using an evaluation procedure of only a given degree of accuracy and with an upper bound on cost. This hypothesis is implicit in the representation of dotted arrows linking the first and second columns in Figure 2-1. We argue that it is not worthwhile to spend more than a certain amount on evaluating the truth of a proposition of low generality nor to demand more than a certain degree of confidence that the correct inference is drawn, other things being equal. (The general level of testing costs that are worth incurring will also depend upon such considerations as the stakes involved in the environment in which the system is to operate. Is it a system to decide when to press the nuclear button or a system to control ice-cream quality in a drug-store?)

The contention that generality of hypothesis and costliness of testing go together can be considered through a discussion of the appropriate test for making a discrimination about standard operating procedures in the design of decision-aiding systems. In particular, consider the low generality hypothesis H: "Partitioning enemy threat possibilities into more than three alternatives is rarely called for in preprogrammed decision analysis for tactical warfare." We could, at least in principle, contemplate controlled experiments to test this hypothesis. But one of the requirements of the classical experimental procedure, namely, the replication of the circumstances of the hypothesis and its negative, would call for setting up experimental wars and observing whether the premise or its converse produced a greater tendency to victory! This procedure would be subject to a problem well-known to social scientists, namely, that experimental conditions are often not replicable.

We could test our hypothesis, H, by prototype testing. At the most costly, military exercises carried out by using three-way partitioning of events and the results could be compared to similar exercises using finer partitioning. It is intuitively clear (and judgments between different procedures must, in the last resort, be made by intuition or scientific common sense) that any routine comparison of performance measures in the two kinds of exercise is unlikely to be instructive. Effects caused by differences in the analytic method will be completely swamped by other effects caused by differences in disturbance factors between military exercises. On the other hand, more informal evaluation of how the two variants worked is likely to be very revealing, but the cost of the test would be obviously prohibitive if this were the only justification of the exercise. If the

cost is shared by using the exercise to test several design discriminations, the problem of compounded effects is magnified. However, if a complete decision-aiding system is installed to test the much more general hypothesis, H2, that "a system of this kind is preferable to the status quo," then the high cost may well be justified.

The next most precise evaluatory procedure we have considered, conceptual simulation, can be said to overcome the difficulty of uncontrolled sources of response variation. It is possible to create a computer-based model of a battle, including all the decision-makers, where disturbance factors are held stable and only the analytical tool used is varied. But we argue that to create a simulation which adequately represents a naval battle, including all the bureaucratic stresses on the commanders, is infeasible; moreover, such an adequate simulation would again be prohibitively expensive. For if we are interested in how commanders prove able to use different analytic techniques for decision making, detailed predictive models for how commanders react are essential; and we all know how elusive it is to model human behavior. These difficulties might be somewhat alleviated by using war-gaming, but even here there are substantial problems in constructing a simulation that properly captures the essentials of real battles.

That there exists an upper limit to the money worth expending on evaluating the truth of an hypothesis such as H is evidenced by observing that the cost of an error in inference is quite low. Suppose we infer that H is true and it turns out to be false; it is very unlikely in this case that a disaster would result from this false inference. Moreover, the cost of redesigning the system to cope with it is slight. So we gain very little (and indeed might lose a lot) by paying a great deal for an accurate inference about an hypothesis like H. Moreover, the lifetime of military systems is quite short. After three years or so, the system may well be redesigned in such a way that a particular component is no longer needed. In other words, high development costs involved in testing an hypothesis like H by an expensive and accurate testing procedure may well not be offset by improved performance in the long run.

In general, of course, false inferences can be very costly. In 1973, the Israelis inferred that the Egyptians would not cross the canal, an almost disastrous miscalculation. In the system testing field, a false inference about hypothesis H2, that a new kind of decision-aiding system should replace the old, could lead to lost battles.

We feel that H might be better tested by using the relatively inexpensive method of clinical observation. We

start with the intuition, based on other experience, that untrained decision-makers find probability distributions with more than three outcomes confusing to work with, and we design our decision-aiding system software to include no more than three outcomes. We then observe in workshop trials a number of cases of the use of our system and informally assess how valuable it is according to criteria selected for the purpose and the judgment of experienced commanders.[2] We also try a few commanders on systems with finer outcome partitioning and see how they progress. On this basis, we build up an opinion about the truth of H. This procedure is not very costly, and neither, it must be admitted, is the inference very precise. But if, as we suggest, there are small penalties for error for highly specific hypotheses of this kind, we do not need to make more accurate inferences.

Similar arguments can be made at other levels of generality; for example, it is worth spending a great deal of effort establishing laws of human behavior since they will have a wide applicability and need to be as accurate as possible. Though costly, classical experimentation is clearly the appropriate way to test such hypotheses. Similarly, the nature and behavior of naval command, still described by a quite general class of hypotheses, is best elicited by some form of experimental observation.

But the designer of a decision-aiding system takes such hypotheses as given; he is much more concerned with considerations at a lower level of generality, namely, the design of system structure and detailed system components. According to the parallelism made above and illustrated in Figure 2-1, this emphasis implies that the kinds of test procedures appropriate for his discriminations are more likely to be clinical observation and intuition than experiments, whether on simulations of the system environment or on the environment itself. However, when the more general system design decisions have to be made (such as hypothesis H2) and acted upon with the commitment of major resources (such as throwing out the old system and installing a new one throughout the Navy), then the more accurate and costly evaluations provided by prototype testing (fleet exercise) are called for.

The scheme of Figure 2-1 can be justly criticized because the correspondence between the columns is not invariable (e.g., clinical observations are sometimes more accurate than simulation, some generalizations about man require only

---

[2]For an example of such an assessment involving three major decision aid options (simple vs. standard vs. complex model) in a specific scenario, each evaluated according to 32 criteria, see Figure 4-1 in Brown et al. (1975).

a limited accuracy in their inference, and some highly specific discriminations require high accuracy). But we believe it to represent a valid general argument, namely, that the method appropriate to evaluate some aspect of a decision-aiding system depends on the generality of that aspect and, in particular, that the more general the statement and the wider validity sought for it, the more is it worth spending money to gain an accurate inference by using a more powerful procedure.

3.3 Sequential Testing

We have noted above that the design process is sequential because a succession of design discriminations from the general to the specific is typically involved. In addition, and closely related to the design sequence, the testing procedure for <u>any given stage</u> in the design sequence is also sequential because a succession of tests is in principle possible. The designer first thinks about a general idea to see if it makes sense. Then he solicits the opinions of potential users, constructs a prototype to a finer specification, and finally implements the final design in full-scale production. Of course, the stages in this process differ slightly depending on the object of design. For example, a particular freeway system would not be preceded by a prototype, but a naval decision-aiding system would be.

The general principle is that between proposing and adopting a design option, one may apply tests of increasing accuracy and cost up to the limit justified by the nature of the design option. One starts with a low validation hurdle (intuitive plausibility) and only if that is cleared is a higher hurdle (e.g., prototype or test market) attempted. In this way, the unnecessary application of the more expensive tests can be avoided.

REFERENCES

Brown, R. V. Research and the Credibility of Estimates. Cambridge, Mass.: Harvard Business School, 1969.

Brown, R. V.; Hoblitzell, C. M.; Peterson, C. R.; and Ulvila, J. W. Decision Analysis as an Element in an Operational Decision Aiding Sytem. Technical Report 74-2. McLean, Va.: Decisions and Designs, Incorporated, September, 1974.

Brown, R. V.; Peterson, C. R.; Shawcross, W. H.; and Ulvila, J. W. Decision Analysis as an Element in an Operational Decision Aiding System (Phase II). Technical Report 75-13. McLean, Va.: Decisions and Designs, Incorporated, November, 1975.

Brown, R. V., and Ulvila, J. W. Selecting Analytic Approaches for Decision Situations: A Matching of Taxonomies. Technical Report 76-10. McLean, Va.: Decisions and Designs, Incorporated, October, 1976.

Chase, Wilton P. Management of Systems Engineering. New York: John Wiley, 1974.

Chinnis, J. O.; Kelly, C. W., III; Minckler, R. D.; and O'Connor, M. F. Single Channel Ground and Airborne Radio System (SINCGARS) Evaluation Model. Technical Report DT/TR 75-2. McLean, Va.: Decisions and Designs, Incorporated, September, 1975.

Cox, D. R. Planning of Experiments. New York: John Wiley, 1958.

Fisher, R. A. Design of Experiments. 1951; rpt. New York: Hafner Press, 1974.

Glegg, G. L. The Design of Design. Cambridge, Eng.: Cambridge Univ. Press, 1969.

Glegg, G. L. The Science of Design. Cambridge, Eng.: Cambridge Univ. Press, 1973.

Hays, M. L.; O'Connor, M. F.; and Peterson, C. R. An Application of Multi-Attribute Utility Theory: Design-to-Cost Evaluation of the U.S. Navy's Electronic Warfare System. Technical Report DT/TR 75-3. McLean, Va.: Decisions and Designs, Incorporated, October, 1975.

Keeney, R. L., and Raiffa, H. Analysis of Decisions with Multiple Objectives. New York: John Wiley, 1976.

Lucas, H. C., Jr. *Toward Creative Systems Design*. New York: Columbia Univ. Press, 1974.

Marples, D. L. "The Decisions of Engineering Design." *Management Transactions* (Institute of Engineering Designers) (1961): 1-16.

Peterson, C. R.; Randall, L. S.; Shawcross, W. H.; and Ulvila, J. W. *Decision Analysis as an Element in an Operational Decision Aiding System (Phase III)*. Technical Report 76-11. McLean, Va.: Decisions and Designs, Incorporated, October, 1976.

Raiffa, H. *Decision Analysis*. Reading, Mass.: Addison-Wesley, 1968.

Watson, S. R., and Brown, R. V. *Issues in the Value of Decision Analysis*. Technical Report 75-9. McLean, Va.: Decisions and Designs, Incorporated, October, 1975.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director Advanced Research Projects Agency Attention: Program Management Office 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Office of Naval Research Attention: Code 455 800 North Quincy Street Arlington, Virginia 22217	3 copies
Defense Documentation Center Attention: DDC-TC Cameron Station Alexandria, Virginia 22314	12 copies
DCASMA Baltimore Office Attention: Mr. K. Gerasim 300 East Joppa Road Towson, Maryland 21204	1 copy
Director Naval Research Laboratory Attention: Code 2627 Washington, D.C. 20375	6 copies
Office of Naval Research Attention: Code 102IP 800 North Quincy Street Arlington, Virginia 22217	6 copies

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense
Research and Engineering (Test and Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Assistant Director (Environmental and Life Sciences)
Office of the Deputy Director of Defense
Research and Engineering (Research and Advanced Technology)
Attention: COL Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
APO San Francisco 96610

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Command and Control Technical Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations (OP-987)
Attention: Dr. Robert G. Smith
Washington, DC 20350

Office of the Chief of Naval Operations (OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology (Code 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 436)
Attention: Dr. Bruce MacDonald
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Naval Personnel Research and Development
Center
Acquisition and Initial Service (Code 310)
Attention: Dr. Leonard P. Kroeker
San Diego, CA 92152

Naval Personnel Research and Development
Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N215)
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation Group
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 5403)
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E621
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson (PERI-OK)
1300 Wilson Boulevard
Arlington, VA 22209

Director, Organizations and Systems
Research Laboratory
U.S. Army Institute for the Behavioral
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Technical Director, U.S. Army Concepts
Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine
Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office (Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development
and Acquisition Programs
Office of the Deputy Chief of Staff for
Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Director, Advanced Systems Division
(AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of
Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and
Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Department of Psychology
Montgomery College
Attention: Dr. Victor Fields
Rockville, MD 20850

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Human Factors Research, Incorporated
Santa Barbara Research Park
Attention: Dr. Robert R. Mackie
6780 Cortona Drive
Goleta, CA 93017

Wharton School
University of Pennsylvania
Attention: Dr. H. L. Morton
Philadelphia, PA 19174

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Analytics, Incorporated
Attention: Mr. William Martin
7405 Colshire Drive, Suite 200
McLean, VA 22101

Eclectech Associates, Incorporated
Attention: Mr. Alan J. Pesch
Post Office Box 179
North Stonington, CT 06359

Department of Electrical and Computer Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

CACI, Incorporated-Federal
Attention: Dr. Bertram Spector
1815 North Fort Myer Drive
Arlington, VA 22209

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. T. Owen Jacobs
P.O. Box 3122
Ft. Leavenworth, KA 66027

Dr. Gary Ireland
500 23rd Street, NW - Apt. B301
Washington, DC 20037

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: (14) DT/TR-77-1	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): (6) TESTING PROCEDURES IN THE DESIGN OF MANAGEMENT SYSTEMS: SOME METHODOLOGICAL REFLECTIONS	5. TYPE OF REPORT & PERIOD COVERED: (9) Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): (10) R. V. Brown, S. R. Watson, Engineering Dept., University of Cambridge	8. CONTRACT OR GRANT NUMBER(s): (15) N00014-76-C-0074, N00014-74-C-0263
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Decisions and Designs, Incorporated, 8400 Westpark Drive, McLean, Virginia 22101	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (12) 32p.
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Advanced Research Projects Agency, 1400 Wilson Boulevard, Arlington, VA 22209	12. REPORT DATE: (11) February 1977
	13. NUMBER OF PAGES: 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

management and information systems
engineering design
statistical sampling
decision theory
scientific method

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The introduction (or modification) of a management system in an organization is often preceded by an effort to gather data from which it can be evaluated. The data may come from some kind of experiment, a conceptual simulation, or some more informal analysis of relevant past experience. This paper discusses how such alternative testing procedures can themselves be evaluated by paying particular attention to analogous testing paradigms in the more established fields of science and engineering. Decision-aiding systems for naval command and control are used as an illustrative case.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

390 664

mt